

# A novel heterogeneous ensemble of extreme learning machines and its soft sensing application

Ma Ning Dong Ze

(Hebei Technology Innovation Center of Simulation & Optimized Control for Power Generation,

North China Electric Power University, Baoding 071003, China)

(School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China)

**Abstract:** To obtain an accurate and robust soft sensor model in dealing with the increasingly complex industrial modeling data, an effective heterogeneous ensemble of extreme learning machines (HEELM) is proposed. Specifically, the kernel extreme learning machine (KELM) and four common extreme learning machine (ELM) models that have different internal activations are contained in the HEELM for enriching the diversity of sub-models. The number of hidden layer nodes of the extreme learning machine is determined by the trial and error method, and the optimal parameters of the kernel extreme learning machine model are determined by cross validation. Moreover, to obtain the best output of the ensemble model, least squares regression is applied to aggregate the outputs of all individual models. Two complex data sets of practical industrial processes are used to test the HEELM performance. The simulation results show that the HEELM has a high prediction accuracy. Compared with the individual ELM models, bagging ELM ensemble model, BP and SVM models, the prediction accuracy of the HEELM model is improved by 4.5% to 8.7%, and the HEELM model can obtain better generalization capability.

**Key words:** soft sensor; extreme learning machine; least squares; ensemble

**DOI:** 10.3969/j.issn.1003-7985.2020.01.006

In many industrial processes, some key process parameters are of great importance to the implementation of control strategies and production plans<sup>[1]</sup>. However, in some situations, due to the technical problems, a high investment cost or measurement delay, it is difficult to obtain these variables using hardware sensors<sup>[2]</sup>. To solve this issue, soft sensing technology has been studied and applied by many scholars in the past decades<sup>[3-5]</sup>.

**Received** 2019-10-12, **Revised** 2019-12-28.

**Biographies:** Ma Ning (1992—), male, Ph. D. candidate; Dong Ze (corresponding author), male, doctor, professor, dongzencepu@163.com.

**Foundation items:** The National Natural Science Foundation of China (No. 71471060), the Natural Science Foundation of Hebei Province (No. E2018502111), Fundamental Research Funds for the Central Universities (No. 2019QN134).

**Citation:** Ma Ning, Dong Ze. A novel heterogeneous ensemble of extreme learning machines and its soft sensing application[J]. Journal of Southeast University (English Edition), 2020, 36(1): 41 – 49. DOI: 10.3969/j.issn.1003-7985.2020.01.006.

Soft sensor modeling methods can be divided into the mechanism modelling method and data-driven modeling method<sup>[6-7]</sup>. The mechanism modeling method has the advantages of strongly explanatory and easily understood, but it also has the disadvantages of a complex model and poor portability, especially for some complicated thermal and chemical processes. The other soft sensor modelling method, namely data-driven method, can be developed through learning the historical data. There are various methods applied to setting up data-driven models, for instance, a support vector machine<sup>[8]</sup>, Gaussian process regression<sup>[9]</sup>, artificial neural networks (ANNs)<sup>[10]</sup>, and so on. Compared with other methods, ANNs show prominent advantages due to their good non-linear mapping and generalization ability. Hence, ANNs have been used in a wide variety of industrial process modeling<sup>[11-12]</sup>.

Actually, the accuracy and stability of soft sensor models are the most important criteria to evaluate the quality of models established. In spite of having strong fitting and generalization capability, ANNs are essentially unstable methods based on the statistical theory. The output of ANNs highly depends on the initial weight and training samples. Previous studies also have shown that the performance of a single neural network model is unstable. The performance of ANNs depends heavily on the model structure, especially for the number of nodes and layers in the hidden layers. With the increase in industrial complexity, the dimensionality and coupling of process data tend to be larger, which undoubtedly increases the difficulty of data-driven modelling methods. Hence, some efforts have been made to increase the generalization and stability capability by scholars through various technical methods, for instance, the ensemble method, regularization approach, and so on. Among the above techniques, the ensemble approach seems to be pretty effective. Hansen et al.<sup>[13]</sup> firstly proposed ANN ensemble in 1990. Many previous studies have confirmed that the neural network ensemble can show better performance for the same issue by aggregating the outputs of some individual neural networks<sup>[14]</sup>. The reason why the ensemble learning model exhibits a high prediction accuracy is that the ensemble method can balance the outputs of multiple individual subnets, weakening the influence of imperfect models.

Although ANN ensemble approaches have a wide application in practice, one important issue should be considered. General neural network models in ANN ensemble use very time-consuming training methods, such as the back propagation (BP) method to train the model, which suffers from some insuperable disadvantage, such as a plenty of adjustable parameters and danger of over-fitting<sup>[15]</sup>. To deal with this difficult problem, a kind of effective ANN model called extreme learning machine (ELM) is selected. Different from other neural network methods, ELM transforms the learning training problem into solving the least squares norm problem of the output weight matrix, which gives it the advantages of avoiding falling into local extremism and having a powerful generalization capability<sup>[16]</sup>. Moreover, many kinds of activation functions can become an ELM inner function, regardless of whether the function is continuous or discontinuous. Due to these advantages, ELM is used to construct an ANN ensemble model in this work. However, the common ELM model uses single type of activation function, which can restrict the performance and robustness of ELM.

To eliminate such restrictions, the heterogeneous ensemble model based on kernel extreme learning machine (KELM) and multiple inner functions of ELMs (HEELM) is developed. In the proposed HEELM ensemble model, five kinds of ELMs (sigmoid activation functions ELM, sin activation functions ELM, radbas activation functions ELM, tribas activation functions ELM and one KELM) are selected as individual models. Meanwhile, to further improve the performance of the ensemble model, least squares regression is used to aggregate the outputs of each signal models. In order to validate HEELM performance, the HEELM is used to establish the soft sensor models of two real-world complex datasets, and simultaneously, unlike other single models. Finally, test results prove that the proposed HEELM has both a good generalization capability and strong robustness.

## 1 Theory and Algorithm

### 1.1 ELM

ELM was firstly proposed by Huang et al.<sup>[17]</sup>, and it has been widely applied in various fields in recent years. The structure of the ELM is given in Fig. 1, where we can recognize that the ELM is a three-layer neural network. Compared with the traditional BP or RBF network, the ELM model has a relatively fast learning speed, the reasons of which lie in two aspects: One is that the biases and input weights of the ELM are randomly assigned, and the other is that the least square approach is applied to calculate the output weights of the ELM. The procedure of the ELM algorithm is exhibited below.

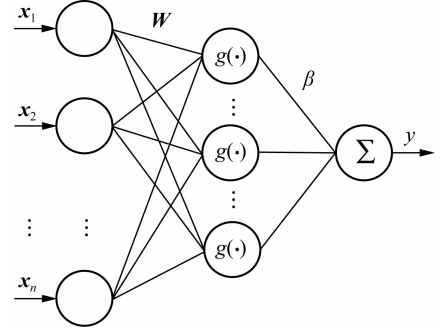


Fig. 1 The structure of ELM

Suppose that there are  $N$  training samples  $(\mathbf{x}_i, \mathbf{t}_i)$ , in which  $\mathbf{x}_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}^T \in \mathbf{R}^n$ ,  $i = 1, 2, \dots, N$ , are the input data and  $\mathbf{t}_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in \mathbf{R}^m$ , are the output data.  $n$  and  $m$  are equal to the number of input layer nodes and output nodes of the ELM, respectively. The following form is the computational expression of ELM,

$$\sum_{i=1}^l \beta_i g_i(\mathbf{x}_j) = \sum_{i=1}^l \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = o_j \quad j = 1, 2, \dots, N \quad (1)$$

where  $\beta_i$  is the output weights, and it connects the  $i$ -th hidden node with the output nodes. Simultaneously,  $\mathbf{w}_i$  represents the input weights, connecting the  $i$ -th hidden node with the input nodes;  $l$  is the number of hidden layer nodes of the model and  $g(\cdot)$  is the activation function. Previous studies show that the output value of the ELM model can be fitted to samples with zero error. Therefore, a derivation equation can be obtained as

$$\sum_{j=1}^l \|\mathbf{o}_j - \mathbf{t}_j\| = 0 \quad j = 1, 2, \dots, N \quad (2)$$

Eq. (1) can be written as

$$\sum_{i=1}^l \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{t}_j \quad j = 1, 2, \dots, N \quad (3)$$

Eq. (3) can be simply written as

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T} \quad (4)$$

where

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \dots & g(\mathbf{w}_l \cdot \mathbf{x}_1 + b_l) \\ \vdots & & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \dots & g(\mathbf{w}_l \cdot \mathbf{x}_N + b_l) \end{bmatrix}_{N \times l} \quad (5)$$

$$\boldsymbol{\beta} = [\beta_1, \beta_2, \dots, \beta_l]^T, \mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_N]^T \quad (6)$$

where  $\mathbf{H}$  is the hidden layer output matrix. In the training process, when  $\mathbf{w}_i$  and  $b_i$  of ELM are generated, the output matrix  $\mathbf{H}$  can be obtained, so that the ELM learning training problem is transformed into the least squares norm problem for solving the output weight, and that is

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^+ \mathbf{T} \quad (7)$$

where  $\mathbf{H}^+$  is the Moore-Penrose generalized inverse of  $\mathbf{H}$ .

Hence, the establishment of the ELM model can be achieved by the following four steps:

**Step 1** Divide the data sets into two parts: the training data set and testing data set.

**Step 2** Randomly assign the input weights and biases and initialize the number of hidden layer nodes.

**Step 3** Obtain the output matrix  $\mathbf{H}$ , and calculate  $\beta$  via training data set.

**Step 4** Use the calculated output weights  $\beta$  to calculate the output value of the model with the testing data set.

Obviously, the learning procedure of the ELM is very fast and easy to implement. Nevertheless, there are still some deficiencies when the ELM is used in practice, which are shown as the following three factors: 1) The performance of a single ELM model tends to be affected by the randomly assigned set input weights and biases. 2) An ELM model is only assigned with one activation function, limiting the robustness of the model to a certain extent. 3) When dealing with very complex large-scale data with high collinearity, one standard ELM always shows poor generalization performance.

## 1.2 Kernel ELM

The kernel ELM (KELM) was proposed by Huang et al.<sup>[18]</sup> based on the analysis of the support vector machine theory, and it is an extension of the extreme learning machine method. The KELM uses Mercer's conditions to define kernel matrix  $\mathbf{\Omega}$  and replaces random matrix  $\mathbf{H}\mathbf{H}^T$  in the ELM with the kernel matrix  $\mathbf{\Omega}$ ,

$$\mathbf{\Omega}_{\text{KELM}} = \mathbf{H}\mathbf{H}^T, \quad \Omega_{i,j} = h(\mathbf{x}_i) \cdot h(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j) \quad (8)$$

According to the above formula, the output of the KELM model is as

$$f(\mathbf{x}) = h(\mathbf{x})\mathbf{H}^T \left( \frac{1}{C}\mathbf{I} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{T} = \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{x}_N) \end{bmatrix} \left( \frac{1}{C}\mathbf{I} + \mathbf{\Omega}_{\text{KELM}} \right)^{-1} \mathbf{T} \quad (9)$$

The KELM method does not need to assign the initial input weights and biases as well as the number of hidden layer nodes. The specific form of the kernel function  $K(\mathbf{x}_i, \mathbf{x}_j)$  is the unique parameter that needs to be adjusted. In this paper, the radial basis function is selected as the kernel function,

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\gamma^2}\right) \quad (10)$$

## 1.3 Least square regression

Least squares regression (LSR) is an effective linear statistical regression modeling method. Assume that the data set consists of an input (independent) variable  $\mathbf{X} \in \mathbf{R}^{n \times m}$  and an output (dependent) variable  $\mathbf{Y} \in \mathbf{R}^{n \times 1}$  and both variables are mean-centered and scaled by the standard deviation. The linear relationship between the input and output variables is expressed in the matrix form as

$$\mathbf{Y} = \mathbf{X} \times \mathbf{W} + \mathbf{E} \quad (11)$$

where  $\mathbf{W}$  is the regression coefficient vector, and  $\mathbf{E}$  is the residual error matrix.

The optimal linear regression relationship between the input and output variables can be estimated by the least squares algorithm, assuming that the optimal linear relationship obtained by least squares is

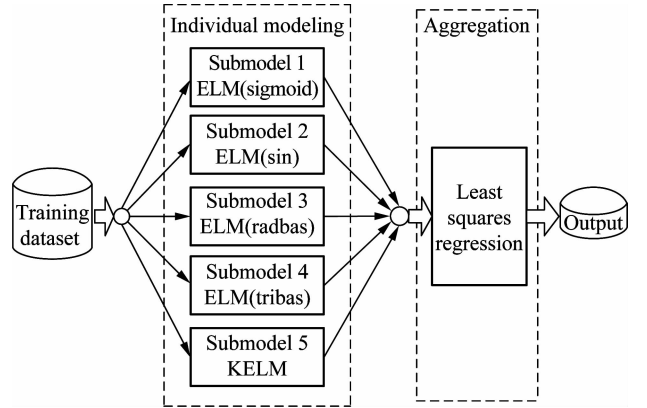
$$\hat{\mathbf{Y}} = \hat{w}_1 x_{i1} + \hat{w}_2 x_{i2} + \dots + \hat{w}_m x_{im} = \mathbf{X}\hat{\mathbf{W}} \quad i = 1, 2, \dots, n \quad (12)$$

$\hat{\mathbf{W}}$  can be calculated as

$$\hat{\mathbf{W}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (13)$$

## 2 Proposed Heterogeneous HEELM Model

To establish a more accurate and stable model for soft sensor modeling, a novel heterogeneous ELM ensemble model called HEELM is developed in this work. The structure diagram of the HEELM is presented in Fig. 2. The proposed HEELM model uses five kinds of ELM to enhance the diversity of the individual model, which also can tackle the problem of noise in training data. As shown in Fig. 2, sigmoid, sin, rabas, tribas function ELM, as well as KELM are applied for the individual model of the HEELM, and the least squares regression method is used as the aggregation strategy to obtain better ensemble outputs. The detailed steps of the HEELM modeling method are described as follows.



**Fig. 2** The structure of the HEELM model

Suppose that the data set is  $D = \{(\mathbf{X}_i, \mathbf{Y}_i) \mid i = 1, 2, \dots, N\}$ , where  $\mathbf{X}_i = [x_{i1}, x_{i2}, \dots, x_{im}] \in \mathbf{R}^m$  represents the input data with  $m$  variables in  $\mathbf{X}_i$ ;  $\mathbf{Y}_i \in \mathbf{R}$  represents output data. Before building the model, the data is divided into three groups: training set  $D_{tr} = \{(\mathbf{X}_t, \mathbf{Y}_t) \mid t = 1, 2, \dots, N_{tr}\}$ , validation set  $D_{va} = \{(\mathbf{X}_v, \mathbf{Y}_v) \mid v = 1, 2, \dots, N_{va}\}$ , and testing set  $D_{te} = \{(\mathbf{X}_{t'}, \mathbf{Y}_{t'}) \mid t' = 1, 2, \dots, N_{te}\}$ ,  $N = N_{te} + N_{va} + N_{tr}$ . The validation set is used to validate the number of hidden layer nodes, the ELM models, and  $C$ ,  $\gamma$  values of KELM.

**Step 1** Preprocess input and output data in the same order of amplitude by the following equations:

$$x_{im} = \frac{x_m^{\max} - x_{im}}{x_m^{\max} - x_m^{\min}} \quad (14)$$

$$Y_i = \frac{Y_i - Y_{\min}}{Y_{\max} - Y_{\min}} \quad (15)$$

where  $x_m^{\max}$ ,  $x_m^{\min}$  are the maximum and minimum of input data. Similarly,  $Y_{\max}$ ,  $Y_{\min}$  are the maximum and minimum of output data, respectively.

**Step 2** Set the input weights and biases of ELM models with sigmoid, sin, rabas, tribas activation functions and build individual models using training set. The KELM model does not need to set input weights and other parameters.  $\gamma$  in the kernel function and regularization coefficient  $C$  are the two parameters in KELM that need to be optimized. In the present study, two parameters are determined by  $k$ -fold cross-validation. Specifically, the training samples are divided into  $k$  groups equally. Then, the  $k - 1$  groups are used to train the KELM model, and the remaining group is applied to the test model. After  $k$  repeated experiments, each group of data can be used as test data in turn. The average of the total test errors is taken as an assessment criterion to evaluate the parameters of the KELM model. Moreover, the most suitable number of hidden nodes of ELM models with sigmoid, sin, rabas, tribas functions is calculated using the trial-and-error approach.

**Step 3** Through the training set  $D_{tr}$ , the outputs of five individual sub models:  $\hat{Y}_{tr}^1$ ,  $\hat{Y}_{tr}^2$ ,  $\hat{Y}_{tr}^3$ ,  $\hat{Y}_{tr}^4$  and  $\hat{Y}_{tr}^5$  are obtained, which are the training output values of five sub-models in Fig. 2, respectively.

**Step 4** Calculate the output of the proposed HEELM model through establishing a regression model between the outputs of each individual model and the expected outputs by the least squares regression technique.

$$\hat{Y}_{tr} = a_1 \hat{Y}_{tr}^1 + a_2 \hat{Y}_{tr}^2 + a_3 \hat{Y}_{tr}^3 + a_4 \hat{Y}_{tr}^4 + a_5 \hat{Y}_{tr}^5 \quad (16)$$

**Step 5** Through the testing set  $D_{te}$ , the outputs of five individual sub models:  $\hat{Y}_{te}^1$ ,  $\hat{Y}_{te}^2$ ,  $\hat{Y}_{te}^3$ ,  $\hat{Y}_{te}^4$  and  $\hat{Y}_{te}^5$  are obtained. Then, the prediction of HEELM model is calculated using the coefficient obtained in Eq. (16).

$$\hat{Y}_{te} = a_1 \hat{Y}_{te}^1 + a_2 \hat{Y}_{te}^2 + a_3 \hat{Y}_{te}^3 + a_4 \hat{Y}_{te}^4 + a_5 \hat{Y}_{te}^5 \quad (17)$$

**Step 6** To accurately evaluate the performance of the proposed HEELM, the root mean square error (RMSE) is used as evaluation criteria. RMSE can be calculated as

$$RMSE = \sqrt{\frac{1}{N_{te}} \sum_{i=1}^{N_{te}} (Y_{te} - \hat{Y}_{te})^2} \quad (18)$$

### 3 Case Studies

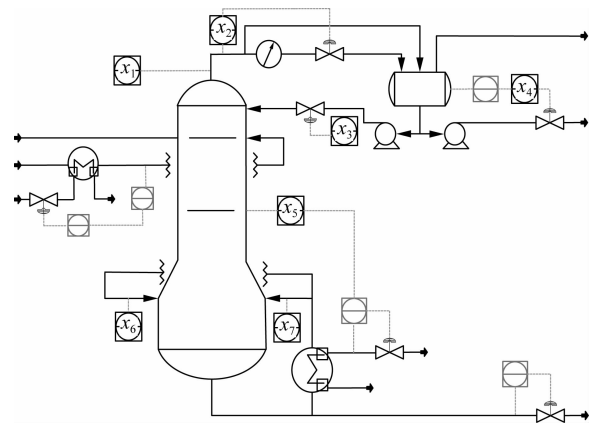
The ensemble model capability is validated using two practical industrial processes: One is the debutanizer and the other is selective catalytic reduction (SCR) flue gas denitration process of the power plant boiler.

#### 3.1 Debutanizer column

The debutanizer column is a part of desulfurization and naphtha splitting plant. Its task is to reduce the concentration of tower bottom butane as much as possible<sup>[19]</sup>. The flowchart of a debutanizer column process is shown in Fig. 3. Usually, the concentration of bottom butane is measured on-line by a gas chromatography analyzer installed on the top of the tower. Since it takes a certain time for the vapor of bottom butane to reach the top of the tower and the analysis process of the gas chromatography analyzer, there is a lag in the on-line measurement of the concentration of bottom butane. So, it is necessary to establish a soft sensor model to estimate the concentration of bottom butane on-line and in real time. There are in total seven variables selected as input variables in the soft sensing model. The only output variable is the concentration of butane in the bottom of the debutanizer. Tab.1 lists the detail description of input variables. There is a total of 2 393 data samples in the debutanizer column process, of which about half are used as training sets, about one-third are test sets and the rest are validation sets. All the data can be downloaded in Ref. [20].

**Tab. 1** Input variables of soft sensor for the debutanizer column

Input variables	Variable description
$x_1$	Top temperature
$x_2$	Top pressure
$x_3$	Reflux flow
$x_4$	Flow to next process
$x_5$	The 6th tray temperature
$x_6$	Bottom temperature
$x_7$	Bottom pressure



**Fig. 3** The flowchart of the debutanizer column

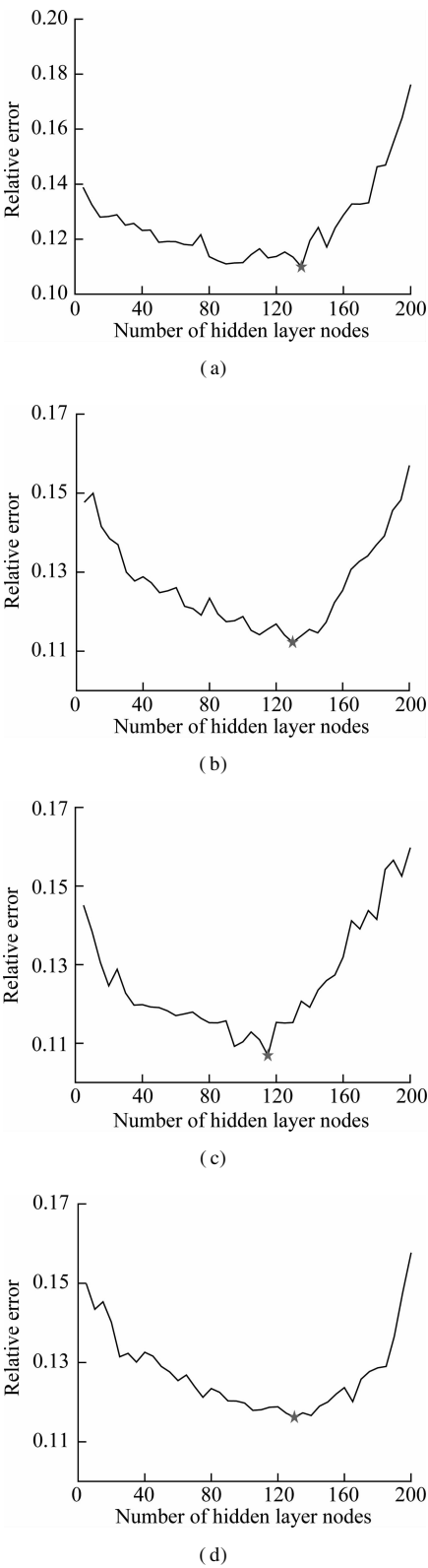
In this study, some kinds of single ELM models include ELM with sigmoid, sin, radbas, tribas activation functions, and the KELM model are built to be compared with the HEELM model. To ensure fair comparison, some parameters for five single models such as the number of hidden layer nodes,  $C$  and  $\gamma$  are firstly selected by the trial-and-error method. Those parameters can be determined when the errors are the smallest within the validation data.

Fig. 4 shows the variation of relative errors of the validation set with the number of hidden layer nodes of ELM models. It can be seen that, for the ELM with sigmoid function, the relative error is the least when the number of nodes is 135. Hence, the number of hidden layer nodes of the individual ELM with the sigmoid function is assigned as 135. Similarly, the numbers of hidden layer nodes of single ELM models with sin, radbas, and tribas inner functions are determined as 115, 130 and 130, respectively. In addition, parameters  $C$  and  $\gamma$  in the KELM model are finally optimized to be  $C = 50$  and  $\gamma = 0.06$ . After determining the optimal parameters of each sub-model, the proposed HEELM can be developed via aggregating the outputs of five individual models using the least squares regression strategy. To enhance the reliability of the simulation experiment, the experiment is repeated 30 times, and the max, min, mean and standard deviation (SD) of RMSE values for the testing dataset are shown in Tab. 2. Bagging the ELM uses five different ELMs as sub-models. Bagging ensemble is a common ensemble technique, and in this study, the Bagging ELM ensemble model is established to make a comparison with the performance of the proposed HEELM.

As seen from Tab. 2, the proposed HEELM model can achieve smaller max, min, and mean of RMSE for the testing dataset than those of the other five individual models and the Bagging ELM model. Fig. 5 displays the variation of RMSE values obtained by the seven models in 30 runs for testing the dataset of debutanizer column. It is clear that, the RMSE value of each ELM models with sigmoid, sin, radbas, tribas activation functions varies from 0.086 9 to 0.110 7 with a large fluctuation. The reason for such a result is that although the optimal number of nodes for each ELM model has been determined, the input weights and bias values of the four ELM models are randomly determined in each simulation experiment, which can lead to the unstable prediction performance of the four models. When the optimum parameters ( $C$ ,  $\gamma$ ) are determined, the KELM model has no other parameters that can be adjusted, so the error results of the KELM model for 30 times are invariable. The RMSE values of the HEELM are low and stable around 0.086 0 without fluctuation.

**Tab. 2** Simulation results of RMSE values for debutanizer column testing dataset

Method	RMSE			
	Max	Min	Mean	SD
ELM( sigmoid)	0.106 7	0.092 7	0.098 6	0.003 7
ELM( sin)	0.103 1	0.089 1	0.097 5	0.003 2
ELM( radbas)	0.110 7	0.092 5	0.099 8	0.004 1
ELM( tribas)	0.104 8	0.086 9	0.095 9	0.003 7
KELM	0.092 6	0.092 6	0.092 6	0
Bagging ELM	0.095 8	0.088 8	0.091 8	0.001 4
HEELM	0.087 7	0.084 3	0.086 1	$8.16 \times 10^{-4}$



**Fig. 4** Variation of relative errors of validation set with the number of nodes of ELMs for the debutanizer column. (a) ELM( sigmoid); (b) ELM(sin); (c) ELM(radbas); (d) ELM(tribas)

Apparently, the HEELM model can achieve much better stability than that of single ELM. In addition, the predictive performance of the KELM model is better than that of other four single ELM models, but not as good as that of the HEELM model. The simulation results of the debu-

tanizer column demonstrate that the proposed HEELM model can achieve better prediction accuracy and model stability.

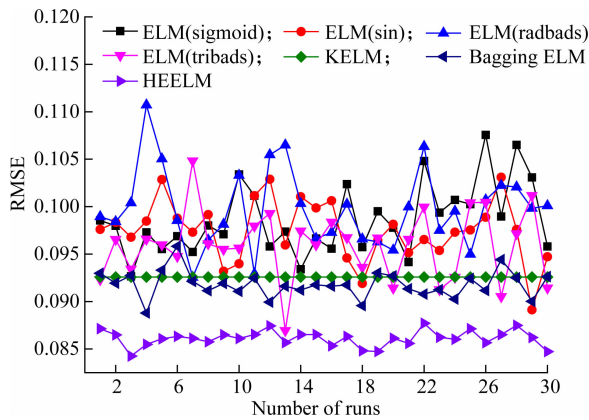


Fig. 5 RMSE values for debutanizer column testing dataset of six models

3.2 SCR flue gas denitration process

SCR flue gas denitrification is a necessary technique in coal-fired power plants for reducing the nitrogen oxides ( $\text{NO}_x$ ). SCR denitrification technique has some salient features such as high denitrification efficiency and simple device structure, so SCR denitrification has attracted much attention and wide application in almost all power plants. The flowchart of SCR flue gas denitration is shown in Fig. 6. The working principle of SCR is that, liquid ammonia reacts with the  $\text{NO}_x$ , and converts  $\text{NO}_x$  to  $\text{N}_2$  and  $\text{H}_2\text{O}$ .

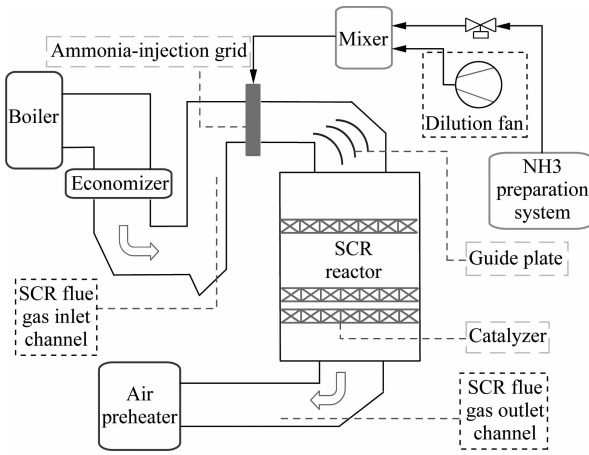


Fig. 6 Schematic diagram of reactor structure in SCR flue gas denitrification system

In this work, 1 000 measurements of a 1 000 MW ultra-supercritical boiler SCR denitrification system boiler operation are obtained from the distributed control system (DCS) database. The sampling interval is 1 min. Based on the basic knowledge of boilers and the engineers' experience<sup>[21]</sup>, six variables are employed as inputs of the SCR model and the only output is the export  $\text{NO}_x$  of the SCR denitrification system. The detailed description of

input variables is listed in Tab. 3. To construct the soft sensor model, 1 000 samples are divided into three parts: 500 samples are used as training sets, 200 samples are the validation sets and the remaining 300 samples are the test sets.

Tab. 3 Input variables of the soft sensor for the SCR flue gas denitration process

Input variables	Variable description
$x_1$	Entrance $\text{NO}_x$ concentration
$x_2$	Inlet gas flow value
$x_3$	Inlet flue gas temperature
$x_4$	Ammonia injection
$x_5$	Unit load
$x_6$	Entrance $\text{O}_2$ concentration

According to the steps of the proposed HEELM approach mentioned above, the number of hidden layer nodes of four common ELM with different activation functions and two parameters ( $C, \gamma$ ) of KELM are firstly determined. Similar to the number of hidden layer nodes determination of debutanizer column simulation in section 3. 1, Fig. 7 presents the relative errors of the validation set with the number of ELM models' nodes. It can be seen from Fig. 7 that, for the SCR flue gas denitration dataset, the most suitable number of hidden layer nodes for four ELM models with sigmoid, sin, radbas, tribas functions is assigned as 85, 90, 100 and 105, respectively. Moreover, according to the cross validation method,  $C$  and  $\gamma$  in the KELM model are assigned to be 50 and 0.1, respectively.

After 30 repeated experiments, the results of the soft model for SCR flue gas denitration case are listed in Tab. 4. Compared with the five single ELM models, BP model and SVM model, it can be clearly seen that the HEELM method can obtain smaller REME values. The SD values of RMSE four common ELM models with sigmoid, sin, radbas, tribas activation functions and BP model are obviously higher than that of the HEELM model, which reveals that the common single ELM model is unstable. Meanwhile, the HEELM model combines the outputs of five ELM models to solve the problem that includes the complex data. Five different kinds of ELM models can realize mutual complementation by least square technique when establishing the soft sensor model. Therefore, the proposed HEELM ensemble model can show the highest accuracy among all the presented models.

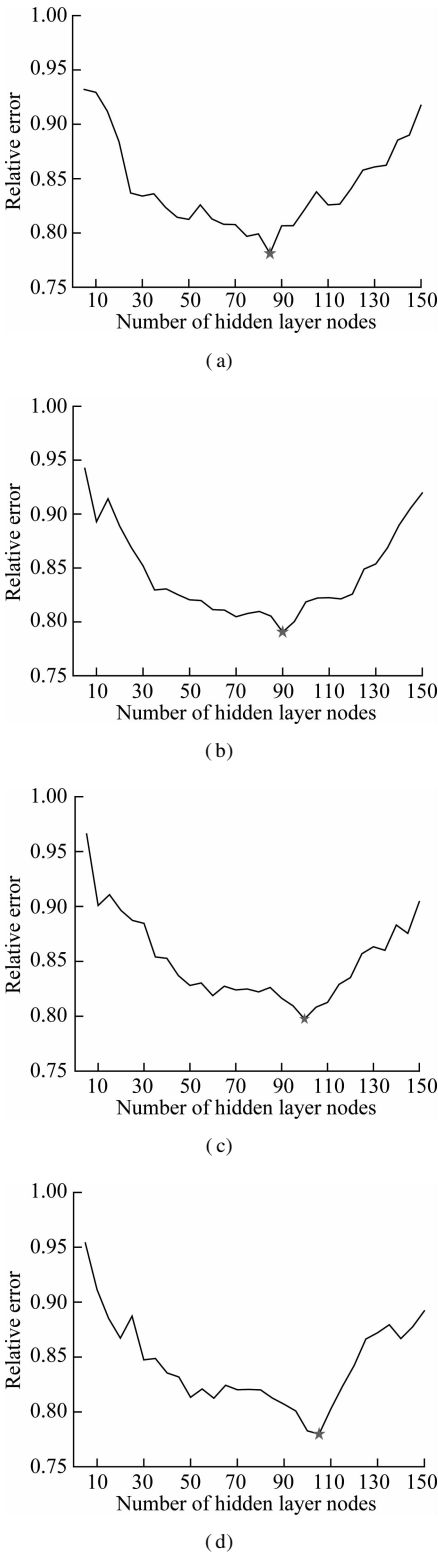
To further show the capability of the HEELM method, a comparison between the predicted results and real data of 300 testing cases is presented in Fig. 8. The red line is the perfect line which shows that predicted values are equal to real values, and the points are the results predicted by the HEELM method. It is easy to see that all of the points distribute closely around the perfect line, which means that the output is the export  $\text{NO}_x$  of SCR which can be predicted with good accuracy by the proposed HEELM for the testing dataset. Moreover, in order to clearly

show the generalization performance of six kinds of ELMs, Fig. 9 presents the variation of RMSE values obtained by the eight models in 30 runs for testing the dataset of SCR flue gas denitration. From Fig. 9, it can be seen that the RMSE values of HEELM are the smallest in all 30 times experiments. Hence, all the simulation re-

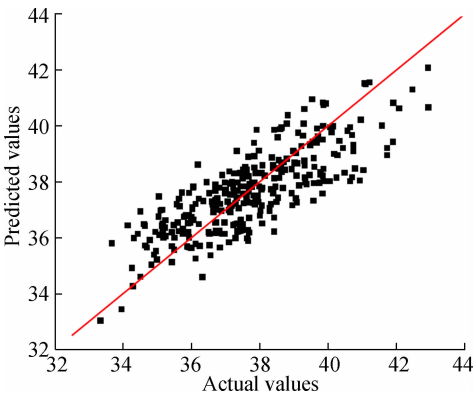
sults of SCR flue gas denitration indicate that the HEELM ensemble model can achieve a high accuracy and good stability.

**Tab. 4** Simulation results of RMSE values for SCR flue gas denitration testing dataset

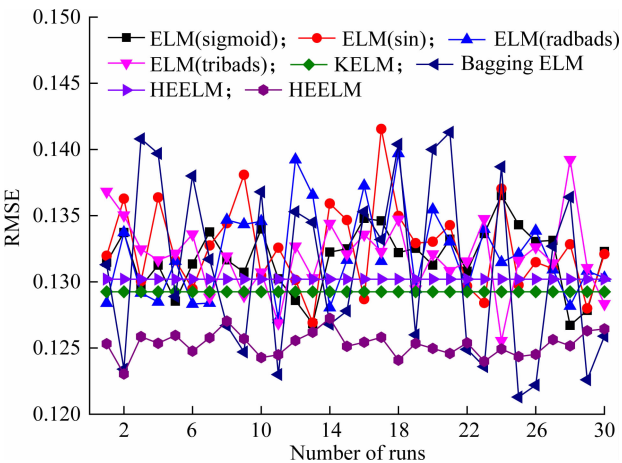
Method	RMSE			
	Max	Min	Mean	SD
ELM( sigmoid)	0. 136 5	0. 126 7	0. 131 8	0. 002 4
ELM( sin)	0. 141 5	0. 126 9	0. 132 6	0. 003 3
ELM( radbas)	0. 139 7	0. 127 2	0. 132 1	0. 003 4
ELM( tribas)	0. 139 2	0. 125 6	0. 131 9	0. 002 8
KELM	0. 129 2	0. 129 2	0. 129 2	0
BP	0. 141 3	0. 121 3	0. 131 1	0. 006 6
SVM	0. 130 2	0. 130 2	0. 130 2	0
HEELM	0. 127 2	0. 123 0	0. 125 3	$9.18 \times 10^{-4}$



**Fig. 7** Schematic diagram of reactor structure in SCR flue gas denitrification system. ( a ) ELM ( sigmoid ); ( b ) ELM ( sin ); ( c ) ELM( radbas ); ( d ) ELM( tribas )



**Fig. 8** Fitting performance of the HEELM model for SCR flue gas denitration testing dataset



**Fig. 9** RMSE values for SCR flue gas denitration testing dataset of eight models

4 Conclusions

1) An advanced approach for soft sensor modeling using a heterogeneous ensemble, namely HEELM, is proposed. Five kinds of ELM algorithms are used for obtaining diversity within the HEELM model in handling complex modeling data. The least square method is used as an effective ensemble technique to enhance the general-

ization ability by ensuring the worst individual model have the least impact on the final output.

2) The generalization performance of the proposed HEELM ensemble model is verified by two real datasets from the debutanizing and the SCR flue gas denitration processes. The simulation results show that the HEELM model can achieve a good performance in generalization accuracy and stability.

3) The modeling performance of the HEELM is also compared with individual ELM models, bagging ELM ensemble model, BP, as well as SVM models, and the results demonstrate that the performance of HEELM is better than that of the other models in the aspects of its predictive accuracy.

4) In future study work, other kinds of aggregating techniques and different neural network ensemble models will be studied and utilized.

## References

- [1] Wang T, Gao H J, Qiu J B. A combined adaptive neural network and nonlinear model predictive control for multi-rate networked industrial process control[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, **27**(2): 416 – 425. DOI: 10.1109/tnnls.2015.2411671.
- [2] Yuan X F, Ge Z Q, Song Z H. Soft sensor model development in multiphase/multimode processes based on Gaussian mixture regression[J]. *Chemometrics and Intelligent Laboratory Systems*, 2014, **138**: 97 – 109. DOI: 10.1016/j.chemolab.2014.07.013.
- [3] Zheng J H, Song Z H. Semisupervised learning for probabilistic partial least squares regression model and soft sensor application[J]. *Journal of Process Control*, 2018, **64**: 123 – 131. DOI: 10.1016/j.jprocont.2018.01.008.
- [4] Popli K, Afacan A, Liu Q, et al. Development of online soft sensors and dynamic fundamental model-based process monitoring for complex sulfide ore flotation[J]. *Minerals Engineering*, 2018, **124**: 10 – 27. DOI: 10.1016/j.mineng.2018.04.006.
- [5] Sun Y M, Wang Y L, Liu X G, et al. A novel Bayesian inference soft sensor for real-time statistic learning modeling for industrial polypropylene melt index prediction[J]. *Journal of Applied Polymer Science*, 2017, **134**(40): 45384. DOI: 10.1002/app.45384.
- [6] Wang F F, Li P F, Mi J C, et al. A refined global reaction mechanism for modeling coal combustion under moderate or intense low-oxygen dilution condition[J]. *Energy*, 2018, **157**: 764 – 777. DOI: 10.1016/j.energy.2018.05.194.
- [7] Bala Subramaniyan A, Pan R, Kuitche J, et al. Quantification of environmental effects on PV module degradation: A physics-based data-driven modeling method[J]. *IEEE Journal of Photovoltaics*, 2018, **8**(5): 1289 – 1296. DOI: 10.1109/jphotov.2018.2850527.
- [8] Wang X X, Hu H L, Jia H Q, et al. SVM-based multi-sensor data fusion for phase concentration measurement in biomass-coal co-combustion[J]. *Review of Scientific Instruments*, 2018, **89**(5): 055106. DOI: 10.1063/1.5007100.
- [9] Ge Z Q. Active probabilistic sample selection for intelligent soft sensing of industrial processes[J]. *Chemometrics and Intelligent Laboratory Systems*, 2016, **151**: 181 – 189. DOI: 10.1016/j.chemolab.2016.01.003.
- [10] Li G Q, Chen B, Chan K C C, et al. Modeling thermal efficiency of a 300 MW coal-fired boiler by online least square fast learning network[J]. *Journal of Chemical Engineering of Japan*, 2018, **51**(1): 100 – 106. DOI: 10.1252/jcej.17we114.
- [11] Mohanraj M, Jayaraj S, Muraleedharan C. Applications of artificial neural networks for refrigeration, air-conditioning and heat pump systems: A review[J]. *Renewable and Sustainable Energy Reviews*, 2012, **16**(2): 1340 – 1358. DOI: 10.1016/j.rser.2011.10.015.
- [12] Naveen Kumar V, Lakshmi Narayana K V. Development of thermistor signal conditioning circuit using artificial neural networks[J]. *IET Science, Measurement & Technology*, 2015, **9**(8): 955 – 961. DOI: 10.1049/iet-smt.2015.0008.
- [13] Hansen L K, Salamon P. Neural network ensembles[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1990, **12**(10): 993 – 1001. DOI: 10.1109/34.58871.
- [14] Pham B T, Shirzadi A, Tien Bui D, et al. A hybrid machine learning ensemble approach based on a Radial Basis Function neural network and Rotation Forest for landslide susceptibility modeling: A case study in the Himalayan area, India[J]. *International Journal of Sediment Research*, 2018, **33**(2): 157 – 170. DOI: 10.1016/j.ijsrc.2017.09.008.
- [15] Samiee K, Iosifidis A, Gabbouj M. On the comparison of random and Hebbian weights for the training of single-hidden layer feedforward neural networks[J]. *Expert Systems with Applications*, 2017, **83**: 177 – 186. DOI: 10.1016/j.eswa.2017.04.025.
- [16] Li G Q, Niu P F, Duan X L, et al. Fast learning network: a novel artificial neural network with a fast learning speed[J]. *Neural Computing and Applications*, 2014, **24**(7/8): 1683 – 1695. DOI: 10.1007/s00521-013-1398-7.
- [17] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications [J]. *Neurocomputing*, 2006, **70**(1/2/3): 489 – 501. DOI: 10.1016/j.neucom.2005.12.126.
- [18] Huang G B, Wang D H, Lan Y. Extreme learning machines: A survey [J]. *International Journal of Machine Learning and Cybernetics*, 2011, **2**(2): 107 – 122. DOI: 10.1007/s13042-011-0019-y.
- [19] Fortuna L, Graziani S, Xibilia M G. Soft sensors for product quality monitoring in debutanizer distillation columns[J]. *Control Engineering Practice*, 2005, **13**(4): 499 – 508. DOI: 10.1016/j.conengprac.2004.04.013.
- [20] Shao W M. Research on adaptive soft sensing modeling method based on local learning [D]. Qingdao: China University of Petroleum, 2016. (in Chinese)
- [21] Wu X, Shen J, Sun S Z, et al. Data-driven disturbance rejection predictive control for SCR denitrification system [J]. *Industrial & Engineering Chemistry Research*, 2016, **55**(20): 5923 – 5930. DOI: 10.1021/acs.iecr.5b03468.



# 一种新型异构集成极端学习机模型及其软测量应用

马 宁 董 泽

(华北电力大学河北省发电过程仿真与优化控制技术创新中心, 保定 071003)

(华北电力大学控制与计算机工程学院, 北京 102206)

**摘要:**为了在日益复杂的工业建模数据基础上建立一个准确、稳定的软测量模型,提出了一种有效的异构集成极端学习机(HEELM)软测量模型.采用4种不同激活函数的极端学习机和核极端学习机模型用以丰富集成模型的多样性.极限学习机的隐含层节点数通过试错法确定,并以交叉验证为准则来获得最优的核极限学习机模型参数.为了获得集成模型的最佳输出,采用最小二乘回归方法对所有单个模型的输出进行集成.通过2组复杂的工业过程数据集验证了HEELM模型具有很好的预测精度.与单独ELM模型、bagging ELM集成模型、BP和SVM模型相比,HEELM模型的预测精度提高了4.5%~8.7%,且HEELM模型具有更好的稳定性.

**关键词:**软测量;极端学习机;最小二乘;集成

**中图分类号:**TK22